Effects of low-pass filtering on dialect and gender perception

A Senior Thesis

Presented in Partial Fulfillment of the Requirements for Graduation with Distinction in Speech and Hearing Science in the Undergraduate Colleges of The Ohio State University

By: Zane T. Smith

The Ohio State University

April 2015

Project Advisors: Dr. Robert A. Fox and Dr. Ewa Jacewicz, Department of Speech and Hearing Science

ACKNOWLEDGMENTS

I would like to thank Dr. Fox and Dr. Jacewicz so much for taking me on and allowing me to participate in such an amazing research experiences. I've learned so much about the fields of linguistics and speech science, how new research questions come to be, and how the process to answer these questions works. This experience has been pivotal in my experience as an undergraduate at Ohio State and will shape my future both in graduate school and as a future educator as well. Thank you for being such role models and putting up with me through it all!

I also want to thank Jill Deatheraze and Makenzie Laase for showing me the ropes and making this long process seem much shorter.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

ABSTRACT

In addition to linguistic (message-related) information, spoken language includes indexical information related to the speaker characteristics (e.g., gender, social status, regional identity). This study is an extension of Jacewicz et al. (2015, *JASA*, 137: 2417-2418) which demonstrated that listeners were quite accurate in making decisions regarding the regional dialect and gender of a speaker when responding to short unprocessed phrases from 40 speakers (male and female) from two different dialects spoken in central Ohio and western North Carolina. However, when the signal was low-pass filtered at 400 Hz, sensitivity to dialect dropped significantly. The current study examined performance on the same phrases when the signal was low-pass filtered at progressively higher cutoff frequencies (500, 700, 900 and 1100 Hz) to determine how a series of progressively higher filters influence listeners' perception of talker dialect and sex (Aim 1). In addition, intelligibility of the filtered speech was assessed to determine the optimal filter for removing the semantic content while retaining most of the indexical information (Aim 2). The stimuli were played to 20 listeners (10 male, 10 female) who identified the sex and dialect of the speaker. Listeners were more sensitive to dialect in response to male speech than to female speech. The male talker advantage was manifested predominantly at the two lowest filter cut-offs of 500 Hz and 700 Hz, whereas dialect sensitivity was greatest for female speech at filter cut-off of 900 Hz. Thus, compared with males, there was a 200-Hz upward shift in improved sensitivity to dialect features for females. Intelligibility results further supported the discrepancy. For male speech, the 700 Hz band is the optimal filter for removing the semantic content while retaining most of the dialect-related cues whereas female speech requires a higher filter, about 900 Hz.

Chapter 1

INTRODUCTION

Regional variation in American English spoken in the United States has developed into a very productive area of research in speech communication. Pronunciation patterns across the country have been explored and documented by means of advanced speech analysis and statistical methods. One particular area that has received considerable attention recently is the perception of those regional pronunciation features. Experiments have been conducted to better understand how listeners build perceptual categories for regional dialects and to identify the acoustic cues that contribute most to their classification.

Presumably, perceptual representation of dialect variation is shaped by experience with regional pronunciation features at both segmental (phonemic) and suprasegmental (prosodic) levels of phonological structure. Socio-phonetic work has primarily explored the segmental variation, mostly in vowels and to some extent in consonants (Labov, Ash, and Boberg, 2006; Purnell et al., 2005). To that end, many focused acoustic studies explored fine-grained phonetic details in pronunciation patterns across different geographic regions (Clopper et al., 2005; Fox and Jacewicz, 2009; Irons, 2007). Data also shows that listeners can make perceptual distinctions between different dialects when presented with a range of dialect-specific acoustic cues in unaltered sentences (Clopper and Pisoni, 2004; 2007). Moreover, they can identify dialects on the basis of acoustic details even in isolated one-syllable words (Jacewicz and Fox, 2012; 2014).

Relatively little is known about contribution of suprasegmental cues to dialect identification. These suprasegmental cues are often termed prosodic cues and include rhythm and intonation patterns, along with other subtle variations in lexical stress, pitch range and speaking rate, including the use of pauses. There is mounting evidence that dialects can differ in the way

they utilize prosody. For American English, the reported dialect variations include differences in the rising pitch accents between Minnesotan and Southern Californian speakers (Arvaniti and Garding, 2007) and differences in the frequency of pitch accent types and phrasal-boundary tone combinations between Midwestern and Southern speakers (Clopper and Smiljanic, 2011). Dialect differences were also found in rhythm patterns. In particular, Southern speakers showed more variable vowel intervals than speakers from other major dialect varieties in the U.S., and both Western and Northern speakers had more variable consonant intervals than speakers from other dialect varieties (Clopper and Smiljanic, 2015). The study also revealed that the New England and Southern dialects displayed the most distinctive temporal patterns across several measures, including articulation rate, vowel and consonant variability, and the duration of pauses. Importantly, dialect-specific articulation rate is maintained across the lifespan, irrespective of aging or developmental influences on habitual speech tempo (Jacewicz et al., 2010).

The current study explores the contribution of prosodic cues to dialect identification by systematically removing the segmental and semantic content from speech using low-pass filtering. Low-pass filtered speech retains lower frequency acoustic energy including the tonal quality of the voice. This preserves prosodic aspects of speech such as pitch range, intonation contour, rhythm, speaking rate, and pauses. In general, segmental information should be eliminated with the low-pass filter cut off at 400 Hz. Progressively higher cut-offs permit more graded contributions from segmental sources.

As a method, low-pass filtering has been used in speech research since early intelligibility studies in the 1940s (French and Steinberg, 1947; Pollack, 1948). More recently, low-pass filtered natural speech was used to study vocal emotions in typical speech communication (see

Scherer, 2003, for a review) across different cultures such as English and Japanese (Kitayama and Ishii, 2002) and in individuals with emotional disorders (McNally et al., 2001). Low-pass filtering at various cut-off levels was also applied in research on affectual infant-directed speech (Burnham et al., 2002; Kitamura and Burnham, 2003; Knoll et al., 2009). Lower filter cut-offs such as 255 Hz and 300 Hz were used to study the contribution of voice quality to the perception of race (African American vs. White) and speaker sex (Lass et al., 1980; Thomas and Reaser, 2004).

To date, there are only a handful of studies that used low-pass filtering to investigate the contribution of prosodic cues to the identification of regional varieties of the same language. Bezooijen and Gooskens (1999) examined identification of four regional dialects of Dutch by native Dutch listeners and five regional varieties of British English by native English listeners. Speech samples were obtained from three speakers for each dialect for a total of 12 Dutch and 15 English speakers. For each speaker, a 15-20 second speech passage was created from a longer stretch of spontaneous conversation, and then low-pass filtered at 350 Hz. The results showed that prosodic cues within the 350-Hz frequency band provided very little dialect information relative to both monotonized (removing intonation) and unaltered speech. The authors concluded that prosodic features contribute very little to the identification of regional varieties both in the Netherlands and in the United Kingdom.

Other similar studies include those done by Frota et al. (2002) and Leyden and van Heuven (2006). They studied the effects of low-pass filtering on Brazilian Portuguese (BP) and European Portuguese (EP) and Scottish dialects of English in Orkney and Shetland, respectively. Frota et al. found that there is a rhythmic distinction between the two varieties of Portuguese in that EP is regarded as more stressed-timed and BP is more syllable-timed. The sentences were

low-pass filtered at 400 Hz and presented with either the original intonation contour or with a flattened intonation. Listeners were able to make distinctions between the two varieties only when the intonation pattern was preserved. It was suggested that intonation is a salient and necessary cue for perceiving the rhythmic differences between EP and BP.

Leyden and van Heuven found that there are melodic differences between the two dialects, stemming most likely from the fact that Shetland has retained more of its previous Scandinavian influences than Orkney. In particular, Shetland has a relatively narrow pitch range whereas Orkney is perceived as more "singing" and melodic.). Listeners distinguished clearly between the two dialects when intonation was preserved but were unable to make distinctions when the intonation information was removed in the monotonized condition. These results correspond to those reported for Portuguese (Frota et al., 2002), showing that the prosodic difference between Orkney and Shetland is detectable only when the intonation cue is provided in addition to cues from temporal organization.

In another study, Szakay (2008) investigated prosodic distinctiveness of two varieties of New Zealand English, Maori English (spoken by people of Maori descent) and Pakeha English (the English spoken by European New Zealanders). There is a difference in rhythmic patterns between Maori English and Pakeha English in that the former appears to be more syllable-timed than the latter, although both varieties of English are formally considered stress-timed. Spontaneous speech from 10 Maori and 10 Pakeha English speakers, males and females, was presented to Maori and Pakeha listeners in several degraded conditions. It was found that those listeners who were highly integrated into Maori culture, and thus more experienced with Maori English, classified a more stressed-timed speaker as Pakeha, and a more syllable-time speaker as

Maori. The results indicate that greater exposure to dialect increases sensitivity to the prosodic features of that dialect.

Most recently, low-pass filtered speech was used to examine perceptual distinctiveness of two regional varieties of American English on the basis of prosodic cues (Jacewicz et al., 2015). In that study, spontaneous speech samples from 20 speakers, males and females, from the Midland dialect in Central Ohio and from the corresponding 20 speakers from the Southern dialect in Western North Carolina were low-pass filtered at 400 Hz and presented for dialect and sex identification to Central Ohio listeners. It was found that prosodic cues within the 400-Hz band provided only limited dialect information relative to unaltered speech and that dialect sensitivity was greater in response to male speakers. The sensitivity to speaker sex was high, showing that listeners had no major difficulty making distinctions between male and female voices. Interestingly, sensitivity to speaker sex was greater in response to Ohio speakers, which suggests that a greater exposure to the Ohio dialect facilitated listeners' ability to identify male and female voices.

This study is an extension of Jacewicz et al. (2015). That study did not examine the contribution of filters with frequency cut-offs higher than 400 Hz. However, a lot of indexical (age, sex, social identity) information can actually lie above 400 Hz, which necessitates the use of higher filters such as 700 Hz or 1000 Hz (Knoll et al., 2009). The current study has two aims. The first aim is to examine how a series of progressively higher filters influence listeners' perception of talker dialect and sex. To achieve this aim, the filter cut-offs were 500, 700, 900 and 1100 Hz, which represents a range of progressively higher filters between the low-information cut-off of 400 Hz and unfiltered (or clear) speech. A filter cut-off of 1200 Hz has

been previously used in the literature as a bridge between lower filters and unfiltered speech (Knoll et al., 2009).

The second aim is to determine the optimal filter for removing the semantic content while retaining most of the indexical information. It is unclear how much of the relevant verbal information remains at each progressively higher frequency cut-off and how much speech intelligibility contributes to the perception of talker dialect and gender. While it can be predicted that intelligibility will be severely impaired with frequency cut-offs at 500 Hz or 700 Hz, it could also be the case that higher intelligibility at 900 Hz and 1100 Hz provides additional verbal cues to dialect identification. It is also possible that dialect cues are distributed differently for female and male speech across speech spectrum and thus different filters will supply different sets of cues for dialect and talker identification. It can be reasonably expected that lower fundamental frequency in male speech will supply more prosodic and spectral cues at the lower filter cut-offs when compared with female speech. This, in turn, may contribute to greater intelligibility of male speech at the lower cut-offs. These possibilities will be examined in both identification and intelligibility tasks.

Chapter 2.

METHODOLOGY

2.1. Participants

Twenty participants (10 male, 10 female) between the ages of 19 and 24 years (M = 21.95, SD = 2.09) served as listeners in this study. Participants were recruited by word of mouth. All participants were current or former undergraduate students at The Ohio State University. Five participants were recruited from the Biological Sciences Scholars Program at Ohio State,

six participants were employees at the Recreational and Physical Activities Center at Ohio State, five were recruited from the Department of Spanish and Portuguese at Ohio State, two were recruited from the Department of Speech and Hearing Science at Ohio State, and two were roommates of the author of this thesis. All participants had lived in Columbus continuously for at least 4 years and either spoke or recognized the Midland dialect of American English that is spoken in Columbus. Two participants spent time between parents in Maryland or Pennsylvania and Central Ohio when growing up. One participant was from Oregon, but has lived in Central Ohio for 5 years. Only one participant had undergone speech therapy as a child. All participants reported normal hearing and no disabilities. Subjects were asked to participate in two separate listening tasks on different days between October 2015 and January 2016.

2.2. Stimulus Material

The stimuli were short sentences and phrases spoken by 40 speakers: 20 from Ohio (OH) and 20 from North Carolina (NC) (10 male and 10 female). These were taken from previous recordings of informal talks collected in Central Ohio and Western North Carolina. The speakers ranged in age from 51 to 65 years and represented the regional variant of the Midland and Southern dialects of American English, respectively.

Each speaker contributed 10 different and unique phrases/sentences (N=400). These phrases were, in turn, separated into 10 different sets of 40 sentences, each comprised of one sentence/phrase from each of the speakers. Care was taken to have a similar number of syllables in each set for each dialect, which ranged from 8.4 to 8.9 syllables/sentence (OH mean = 8.45 syll/sent, NC mean = 8.86 syll/sent). Mean duration for OH sentences was 1791.66 ms and for NC sentences was 2063.22 ms. These duration differences reflect dialect-specific differences in

articulation rate, which is greater for OH than for NC (Jacewicz et al., 2009). The sentences within each of the 10 sets were randomized separately.

Five experimental conditions were created. There were four conditions with sentences low-pass filtered at 500, 700, 900 and 1100 Hz, and one condition with the original unprocessed utterances (clear speech). Equiripple low-pass filters were used with stopband frequencies 50 Hz higher in each case, which provided very sharp attenuation slopes. The experimental conditions are summarized in Table 2.1.

Table 2.1. Experimental stimulus conditions.

Condition	Passband Frequency	Stopband Frequency
500 Hz Lowpass	500 Hz	550 Hz
700 Hz Lowpass	700 Hz	750 Hz
900 Hz Lowpass	900 Hz	950 Hz
1100 Hz Lowpass	1100 Hz	1150 Hz
Original speech	none	none

Figures 2.1-2.5 Display spectrograms for the utterance “Now I don’t wanna be cruel” as clear speech, 1100 Hz, 900 Hz, 700 Hz, and 500 Hz respectively

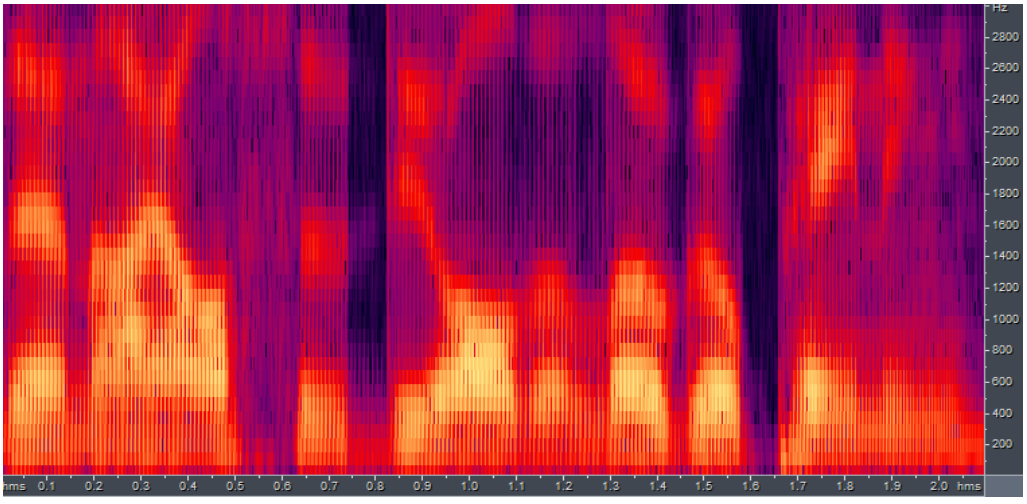

Figure 2.1. “Now I don’t wanna be cruel” as a clear speech waveform

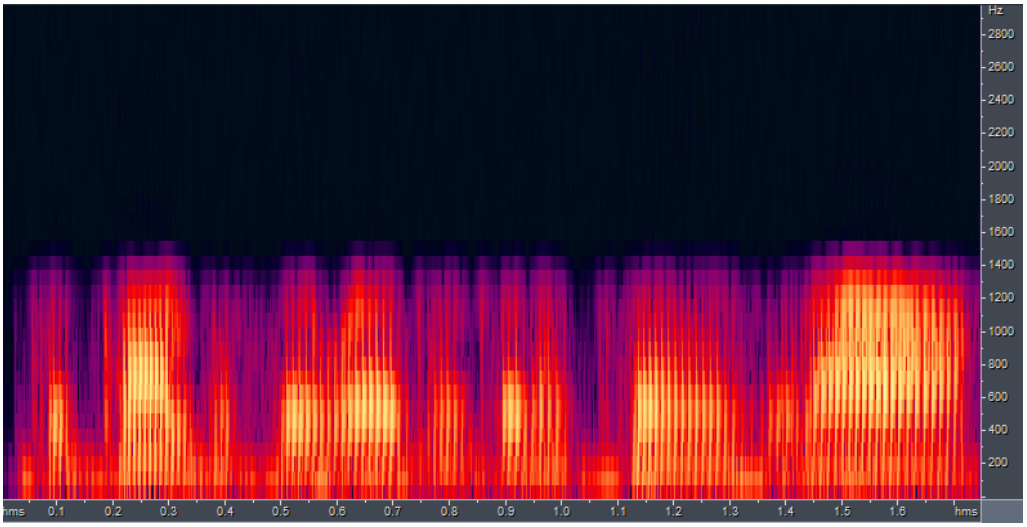

Figure 2.2 "Now I don't wanna be cruel" filtered at 1100 Hz.

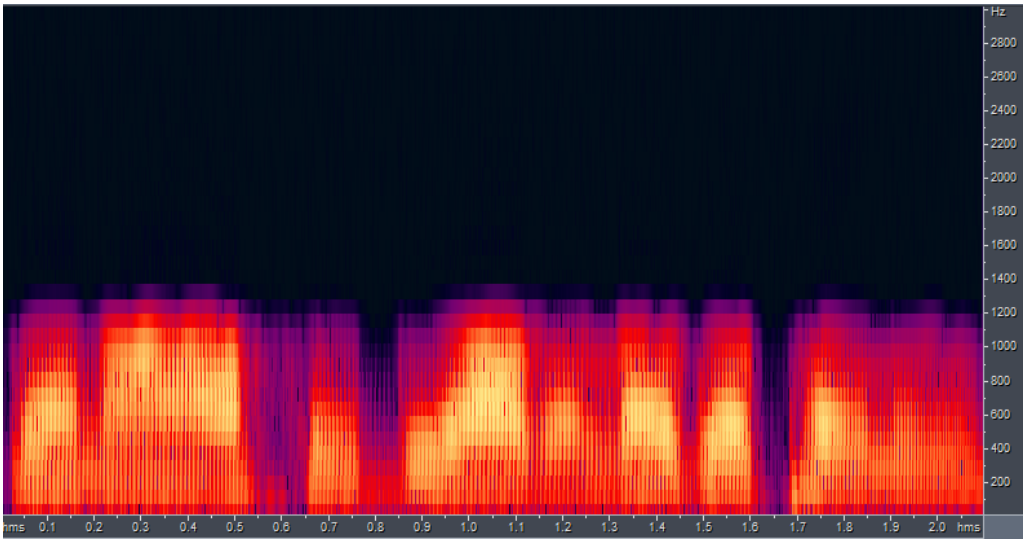

Figure 2.3. "Now I don't wanna be cruel" filtered at 900 Hz.

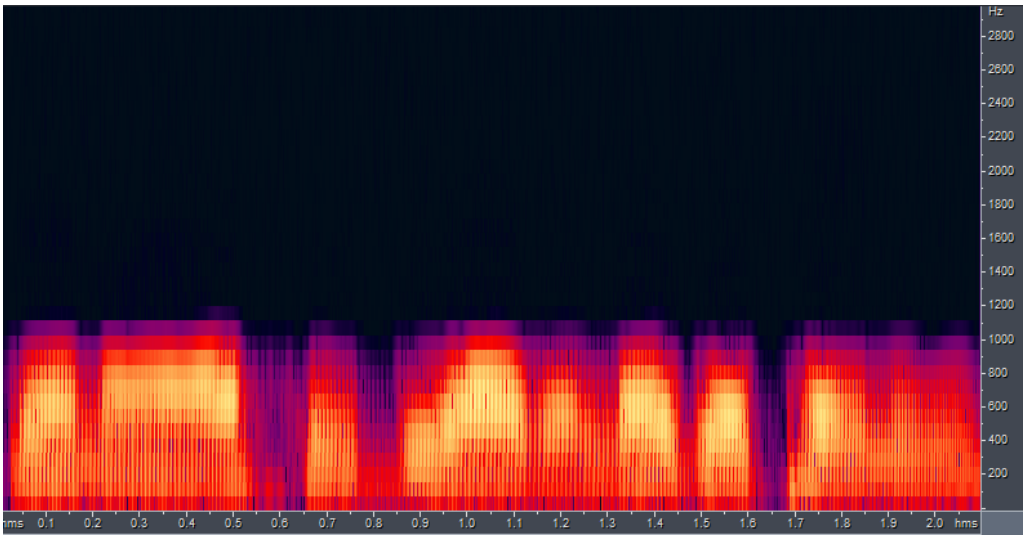

Figure 2.4. "Now I don't wanna be cruel" filtered at 700 Hz.

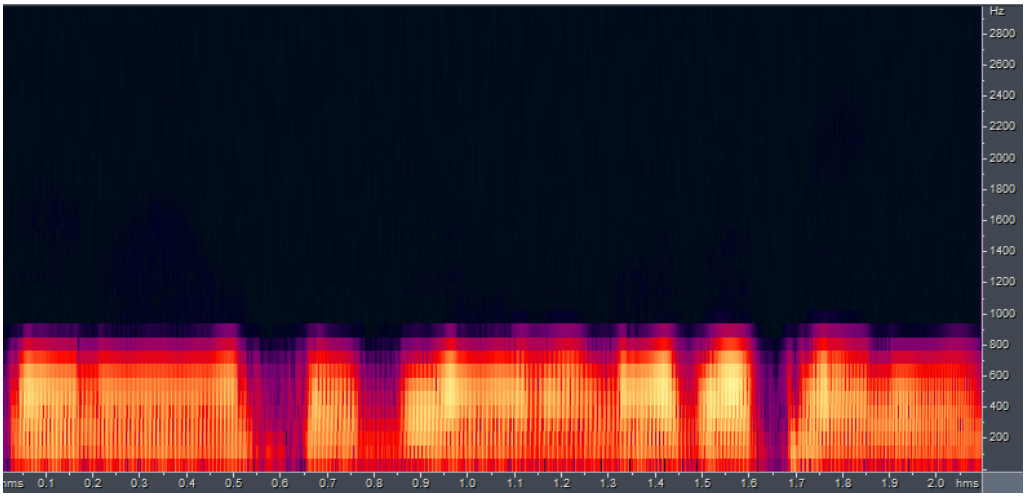

Figure 2.5. "Now I don't wanna be cruel" filtered at 500 Hz.

For each individual listener, two of the 10 stimulus sets were randomly assigned to each of these five conditions so that each listener heard 80 unique sentences/phrases (40 OH, 40 NC) in each of the five conditions. The presentation order of these sentences was also pseudorandomly ordered such that listeners heard sentences in each of the five conditions before

these conditions were repeated. Again, these randomizations were done for each separate listener such that no listener received the same sentence sets in the same conditions in the same order.

2.3. Procedure

There were two listening tasks. In the identification task, participants were asked to identify the sex and dialect of the speakers. Each listener heard one utterance at a time over Sennheiser 640 headphones in a sound attenuating booth. After hearing an utterance, the participants indicated if they thought the speaker was from Central Ohio or North Carolina, male or female. They made their selection on a computer by clicking (using a mouse) on one of four response boxes displayed on the computer monitor in front of them as shown in Figure 2.1.

Figure 2.6. A screen shot of response boxes used by the participant during the identification task to indicate geographic region and sex of the speaker.

In a separate intelligibility task, participants were asked to write down what they heard under the five experimental conditions (four low-pass filtered and one clear speech) from the same 40 speakers. Each listener heard each utterance over Sennheiser 640 headphones in a sound attenuating booth. After hearing an utterance, the participants typed what they heard in the text box (see Figure 2.2) and then clicked "OK" once they were satisfied with the response.

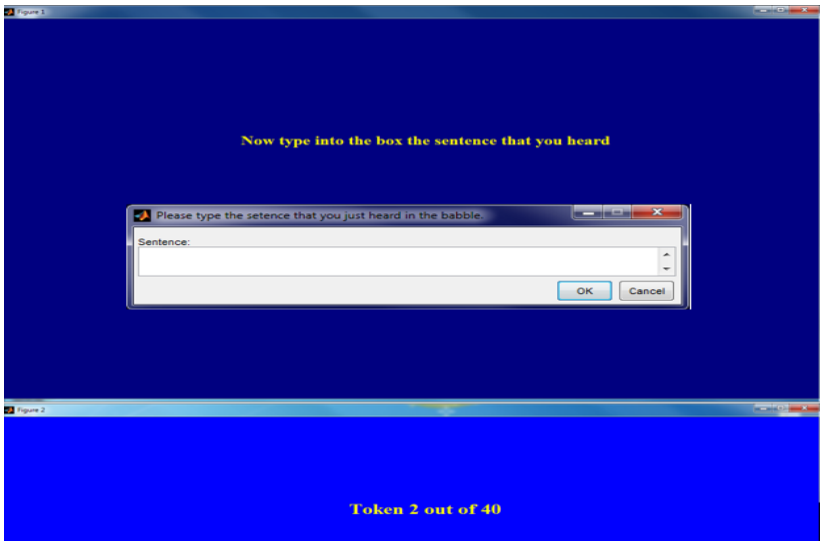

Figure 2.7. A screen shot of the text box used by the participant in the intelligibility task to report what words they heard.

At the first session, each participant was also asked to fill out a background questionnaire that contained questions about his/her speech, language, dialectal, and educational background (See Appendix). This experiment was conducted under a protocol approved by the Institutional Review Board at Ohio State.

For both tasks (i.e., identification and intelligibility), the experimenter verbally explained to the participants seated in the sound attenuating booth that they would be listening to many utterances spoken by male and female speakers from both dialectal regions represented. They

were asked to listen carefully and follow the instructions on the screen to complete the task, depending on which task was being presented. The experimenter then left the booth. Prior to each task, the listeners were also provided with a practice set of 20 sentences with selected low-pass filter levels to ensure they understood the instructions and to verify that the presentation level was comfortable. Participants were allowed to ask questions, express concerns, or take breaks between the blocks. If a participant was unsure of an answer, he/she was instructed to make their best guess. The order of participation in the two tasks was counterbalanced. In particular, 5 males and 5 females started with the Intelligibility Task and 5 males and 5 females started with the Identification Task. Participants were asked to come back at least 2 days later to complete the second task.

The Intelligibity Task was completed in about 2 hours and participants were compensated $15 for their time. The Identification Task was completed in 30-45 minutes and participants were compensated $10 for their time.

Chapter 3.

RESULTS

3.1. Identification task

Listener responses for the identification task were analyzed using Signal Detection Theory (SDT) (Green & Swets, 1966; Macmillan & Creelman, 2005), followed by analysis of variance (ANOVA) and t-tests. Unlike percent correct accuracy scores, SDT is a preferred statistical approach in analyzing listener responses under different degrees of stimulus uncertainty because it allows for the separation of sensitivity and bias (Lynn & Barrett, 2014). In

this analysis, the correct categorization of an OH talker was a hit and the categorization of a NC talker as an OH talker was a false alarm. Nonparametric measures of sensitivity (A') (Snodgrass & Corwin, 1988) and bias (B''_D) (Donaldson, 1992) were used.

3.1.1. Dialect identification

Using IBM SPSS Statistics v. 21 (2012), a two-way repeated-measures ANOVA with the within-subject factors talker sex and low-pass filter level (henceforth LP-level) was used to analyze dialect sensitivity data (A'). A' is a measure whose values range from 0.0 to 1.0. There was a significant main effect of talker sex: Listeners were more sensitive to dialect when responding to male talkers than female talkers, [$F(1, 19) = 15.16$, $p = .001$, $\eta_p^2 = .444$]. Dialect sensitivity as a function of talker sex is illustrated in Figure 3.1.

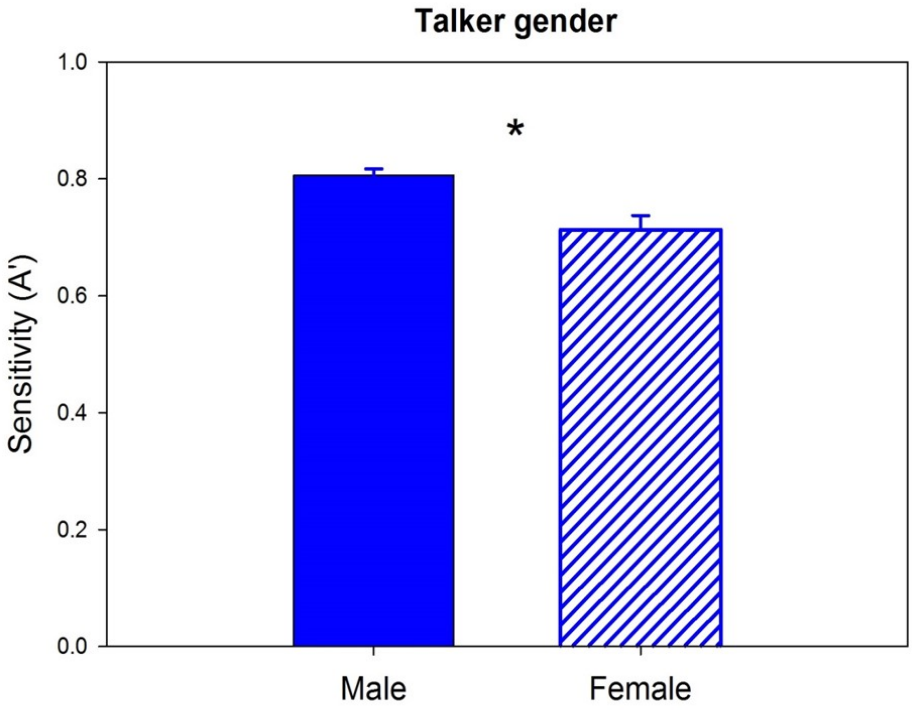

Figure 3.1. Dialect sensitivity by talker sex.

There was also a significant main effect of LP-level: Listeners were more sensitive to dialect at 900 and 1100 Hz cut-offs than at 500 and 700 Hz cut-offs. [$F(4, 76) = 25.87$, $p < .001$, $\eta_p^2 = .577$]. Although the interaction between LP-level and talker sex only approached significance, [$F(4,76) = 2.49$, $p = .05$, $\eta_p^2 = .116$], there was a pattern suggesting that sensitivity to dialect differed between male and female talkers as a function of LP-filter. As shown in Figure 3.2, listeners were more sensitive to male than to female talkers at lower frequency cut-offs (500 Hz and 700 Hz) as well as at 1100 Hz, whereas the talker sex-related differences were reduced at 900 Hz and in the clear speech condition. Subsequent post hoc analyses using paired *t*-tests indicated that listeners were significantly more sensitive to dialect at 700 Hz relative to 500 Hz for both male talkers [$t(19) = -2.96$, $p = .008$] and female talkers [$t(19) = -4.97$, $p = .001$]. Likewise, they were significantly more sensitive to dialect in clear (original) speech relative to the highest filter cut-off at 1100 Hz for both male talkers [$t(19) = -4.97$, $p < .001$] and female talkers [$t(19) = -6.63$, $p < .001$]. However, differences between sensitivity to male and female talkers showed up for the 700 Hz – 900 Hz pair. In particular, sensitivity to male talkers at 900 Hz did not significantly improve relative to that at 700 Hz [$t(19) = -.45$, $p = .658$] whereas the improvement was significant for female talkers [$t(19) = -2.09$, $p < .05$].

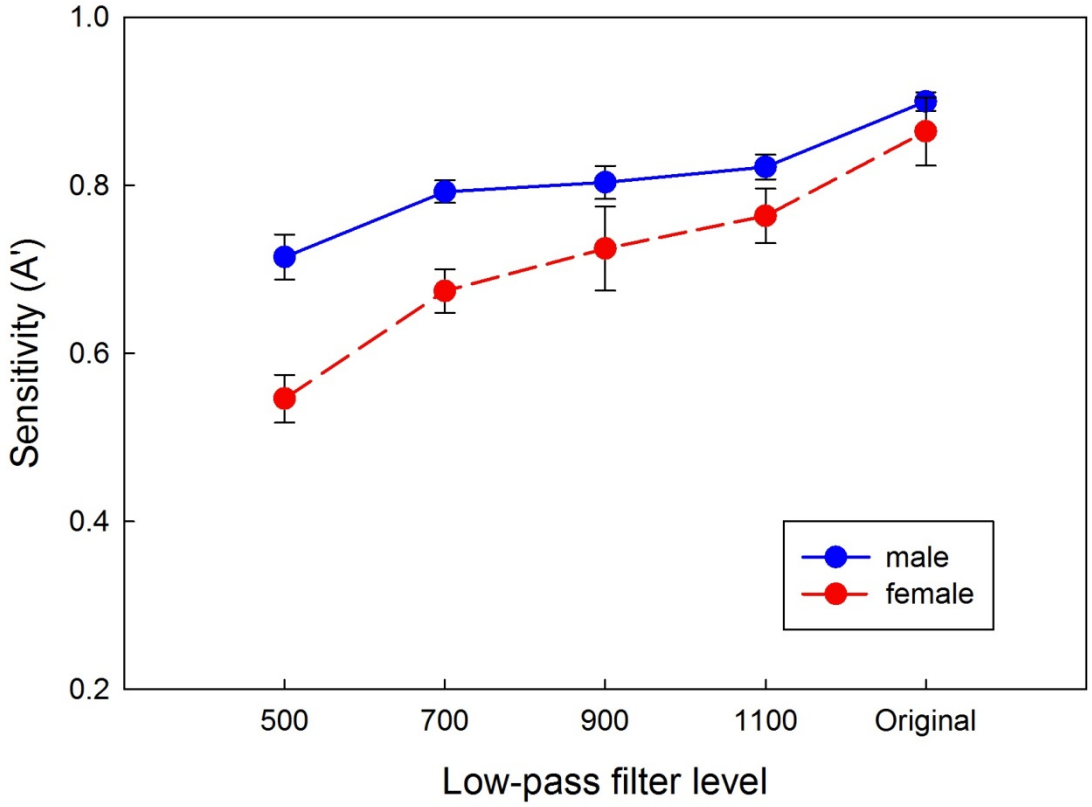

Figure 3.2. Dialect sensitivity to male and female talkers across the experimental conditions (four LP-levels and the original clear speech).

A second repeated-measures ANOVA with the within-subject factors talker sex and LP-level was used to assess possible differences in the response bias (B''_D). The contribution of bias to making decisions about talker dialect reflects how liberal or conservative listeners are under uncertainty (Lynn & Barrett, 2014) and is a function of where each listener places his/her criterion for responding "target." That is, in case of doubt, a conservative listener tends to respond that the talker was from NC (i.e., will choose a foil rather than a target), showing a positive bias. For the B''_D measure, values lie between –1 and +1. A zero value indicates no bias. Negative values are associated with a liberal bias while positive values are associated with a conservative bias.

There were no significant main effects. However, there was a significant interaction between LP-level and talker sex, [$F(4, 76) = 7.38$, $p < .001$, $\eta_p^2 = .280$]. The interaction, displayed in Figure 3.3, was explored using one-sample *t*-tests as post hocs. The analyses showed that, when making decision about talker dialect, listeners were biased toward NC when responding to male talkers at 1100 Hz [$t(19) = 3.42$, $p = .003$] and to female talkers in original (clear) speech condition [$t(19) = 3.99$, $p = .001$]. There were no significant differences from zero ("no bias") between male and female talkers at the remaining LP-levels, indicating that listeners were not biased when responding to speech at filter cut-offs lower than 1100 Hz.

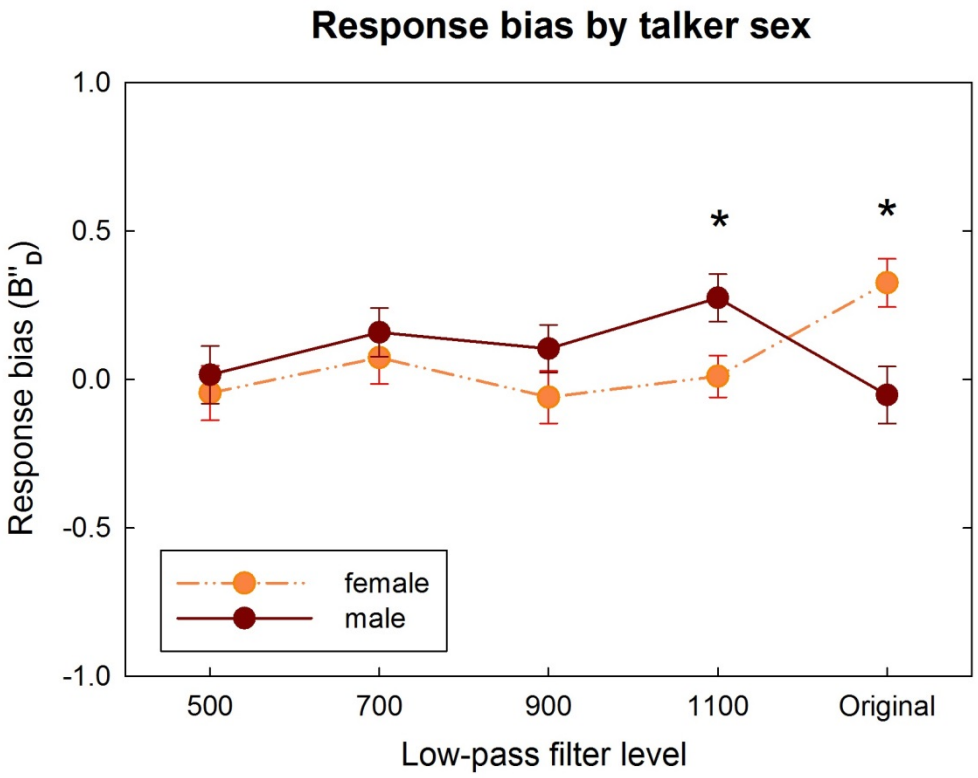

Figure 3.3. Response bias as a function of LP-level and talker sex.

3.1.2. Identification of talker sex

Sensitivity (*A'*) to talker sex was analyzed using a repeated-measures ANOVA with the within-subject factors dialect and LP-level. As shown in Figure 3.4, sensitivity to talker sex was already high at the lowest frequency cut-off of 500 Hz, indicating that listeners had no major difficulties making distinction between male and female voices. There was a significant main effect of dialect: Listeners were more sensitive to talker sex when responding to OH talkers than to NC talkers, [$F(1, 19) = 6.1$, $p = .023$]. There was also a significant main effect of LP-level: Listeners were more sensitive to talker sex when responding to OH talkers at lower cut-offs but dialect did not matter at either 1100 Hz or for unfiltered speech, [$F(4, 76 = 9.62$, $p < .001$].

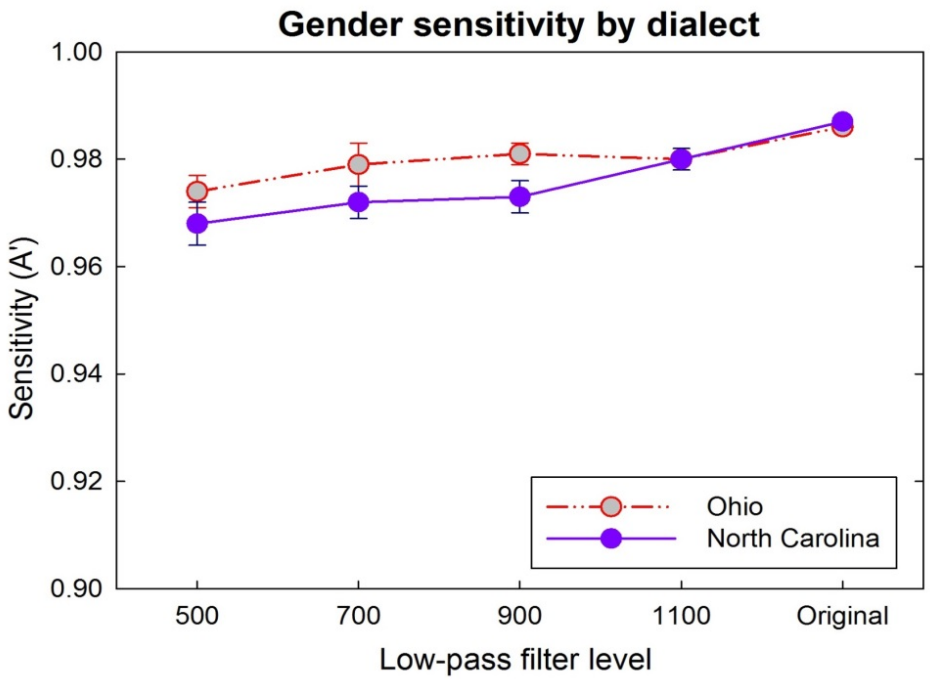

Figure 3.4. Sensitivity to talker sex as a function of dialect and LP-level.

3.2. Intelligibility task

The digitally recorded responses were scored on the basis of keywords. A scoring system was developed for this task. Words with added or deleted morphemes were counted as incorrect and those containing spelling errors were counted as correct. There were 2-3 keywords for each utterance (see the Appendix). Raw scores for each participant were first converted to percent correct and then to rationalized arcsine units (RAU) (Studebaker, 1985) to ensure valid assessment of differences across the entire range of the scale after normalizing for ceiling and floor effects.

A three-way repeated-measures ANOVA with the within-subjects factors dialect, talker sex and LP-level was used to analyze the RAU values. The main effect of talker sex was significant [$F(1, 76) = 16.72$, $p = .001$, $\eta_p^2 = .468$], showing that male talkers were significantly more intelligible ($M = 53.87$ RAU) than female talkers ($M = 49.87$ RAU). The main effect of LP-level was highly significant [$F(4, 19) = 1076.25$, $p < .001$, $\eta_p^2 = .983$]. As expected, intelligibility was low at the lowest LP-level (500 Hz) and progressively increased with each higher frequency cut-off as shown in Figure 3.5. However, as can be seen, the highest frequency cut-off of 1100 Hz still did not deliver a sufficient amount of verbal information to approximate the intelligibility of clear (original and unfiltered) speech. The main effect of dialect was not significant. There was one significant interaction, that between LP-level and talker sex [$F(4, 76) = 7.51$, $p < .001$, $\eta_p^2 = .995$]. The interaction is illustrated in Figure 3.6.

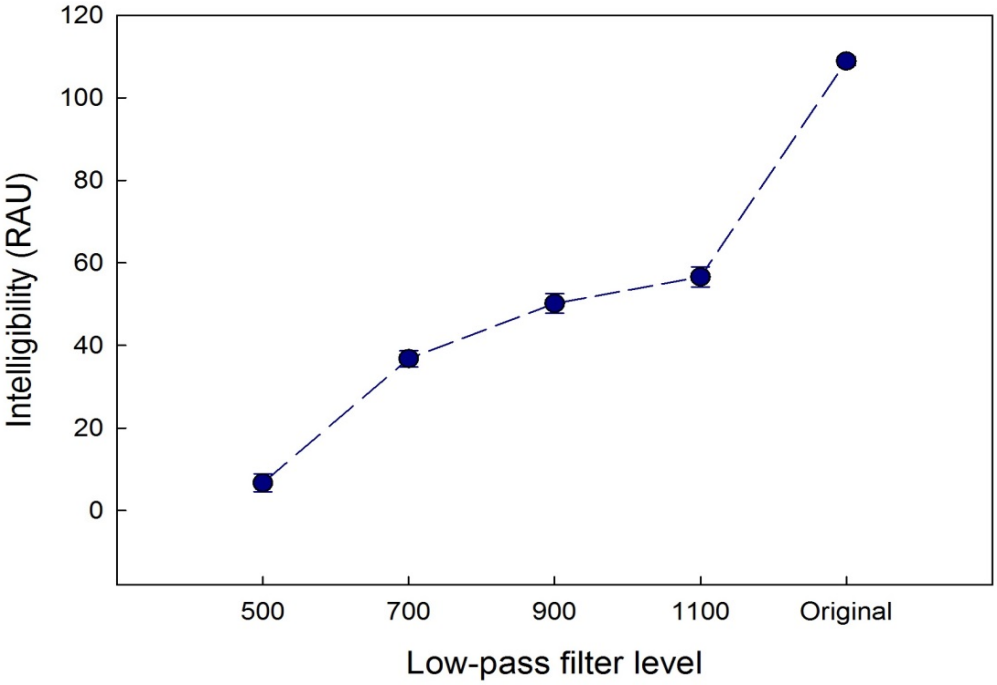

Figure 3.5. Intelligibility by LP-level.

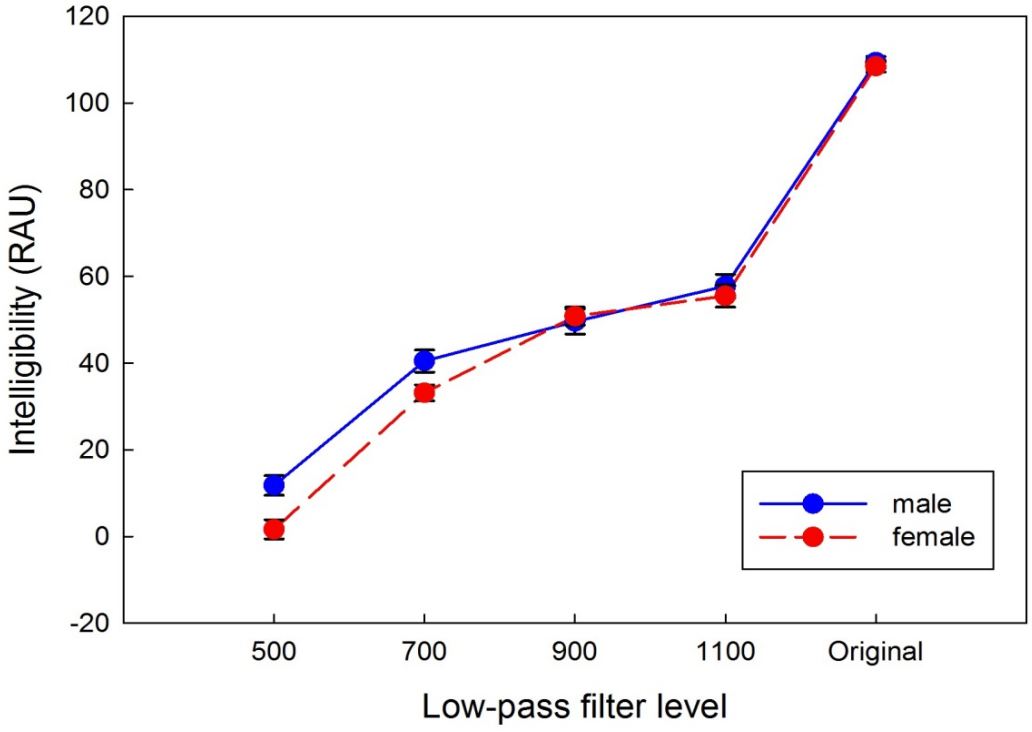

Figure 3.6. Intelligibility as a function of talker sex and LP-level.

The significant interaction between LP-level and talker sex was subsequently explored with paired *t*-tests as post hoc analyses. Significance was considered at a Bonferroni-adjusted level $\alpha = .004$. The analyses showed that intelligibility of male talkers significantly improved with each higher LP-level ($p < .001$ for all pairwise comparisons). However, this was not true for female talkers, whose intelligibility significantly improved with all but one LP-level comparison, that between 900 Hz and 1100 Hz. That is, intelligibility of female talkers at 1100 Hz filter cut-off was not significantly greater than at 900 Hz. Pairwise comparisons between male and female talkers for each LP-level also showed that male talkers were significantly more intelligible than female talkers only at 500 Hz ($p < .001$). Thus, the significant interaction arose because intelligibility of female talkers did not improve at 1100 Hz and there was a significant difference between intelligibility of male and female speech at 500 Hz.

Chapter 4.

SUMMARY AND DISCUSSION

The current research explored the contribution of prosodic cues to dialect identification. The study had two aims. The first aim was to examine how a series of progressively higher filters influence listeners' perception of talker dialect and talker sex. The particular filter cut-offs used were 500, 700, 900 and 1100 Hz. The second aim was to determine the optimal filter for removing the semantic (verbal) content while retaining most of the indexical information related to dialect and talker sex.

4.1. Summary of findings for dialect identification (Aim 1)

The major finding was that sensitivity to dialect features differed as a function of talker sex. In particular, listeners were more sensitive to dialect in response to male speech than to female speech. While this effect was evident across all experimental conditions including both filtered and unfiltered speech, the male talker advantage was manifested predominantly at the two lowest filter cut-offs of 500 Hz and 700 Hz. Also, while sensitivity to dialect features in filtered speech did not substantially improve with higher cut-offs of 900 Hz and 1100 Hz for males, it continued to improve for females. Specifically, dialect sensitivity was greatest for female speech at filter cut-off of 900 Hz. Thus, compared with males, there was a 200-Hz upward shift in improved sensitivity to dialect features for females. No further improvement was found at the highest filter cut-off of 1100 Hz for either female or male talkers. As can be expected, listeners were most sensitive to dialect features in unfiltered (clear) speech, when all original cues in speech signal were preserved.

Another important finding was that listeners were significantly biased in making decisions about talker dialect only when the speech provided more acoustic cues to dialect identification. Again, there were differences related to talker sex. Listeners showed a conservative bias toward NC when responding to female talkers in unfiltered speech and to male talkers at the highest frequency cut-off of 1100 Hz in filtered speech. Listeners were not biased toward either dialect in the remaining conditions. This curious result may indicate that, if uncertain, listeners become more biased toward their non-native dialect (NC) in those situations when listening effort is reduced. However, this explanation is plausible for female speech and only to some extent for male, given that listeners were unbiased when responding to male talkers in the unfiltered (clear) condition when all cues in male speech were available.

As expected, listeners were highly sensitive to talker sex already at the lowest filter cut-off of 500 Hz (mean A' = .971). However, sensitivity still significantly improved with each higher filter, ultimately reaching almost ceiling in the clear speech condition (mean A' = .987). As was observed with a different group of listeners before (Jacewicz et al., 2015), sensitivity to talker sex was significantly greater in response to OH talkers, at least for the three lowest frequency cut-offs. This indicates that when presented with impoverished speech signal, listeners are able to benefit more from acoustic cues to talker sex when they share the same dialect with the talkers.

4.2. Summary of findings for speech intelligibility (Aim 2)

The results show that, overall, each higher filter provided more semantic content but the unfiltered (clear) speech provided the optimal intelligibility cues. For the filtered speech, the intelligibility significantly improvement with each increased LP-level for male talkers. However, intelligibility of female talkers did not significantly improve at filter cut-off of 1100 Hz relative to 900 Hz. Also, there was a male talker advantage at the lowest filter cut-off, that of 500 Hz, in that males were significantly more intelligible than females.

These results indicate that listeners were mostly unable to benefit from verbal information at the lowest LP-level of 500 Hz and, at this filter cut-off, male talkers were somewhat more intelligible than female talkers. The talker sex-related differences were eliminated at the filter cut-off of 700 Hz, however, intelligibility remained well below the chance level (M = 36.77 RAU). Despite the improvement with each higher filter, intelligibility benefit at the highest filter cut-off of 1100 Hz (M = 56.59 RAU) still did not approximate that of clear speech (M = 108.88 RAU).

Clearly, there is a discrepancy between intelligibility and dialect sensitivity (A') at 700 Hz and 900 Hz related to talker sex. First, male talkers were not significantly more intelligible than female talkers at the filter cut-off of 700 Hz (see Figure 3.6), whereas dialect sensitivity was greater for males than for females (see Figure 3.2). Second, intelligibility of male speech significantly improved at 900 Hz relative to 700 Hz but there was no corresponding improvement in dialect sensitivity for males. This indicates that listeners' judgments about the dialect of male talkers did not change when more verbal cues became available at 900 Hz (and mean intelligibility approached 50 RAU). We can thus conclude that listeners made their dialect decisions on the basis of prosodic cues in the frequency band of 700 Hz rather than on the basis of verbal cues (M = 40.45 RAU). It is plausible that, for male speech, the 700 Hz band is the optimal filter for removing the semantic content while retaining most of the dialect-related indexical information.

A different scenario arose for female talkers. In particular, there was a correspondence between a significant improvement in intelligibility and a significant improvement in dialect sensitivity for females at 900 Hz relative to 700 Hz. However, neither dialect sensitivity nor intelligibility significantly improved at 1100 Hz. This indicates that listeners' judgments about dialect of female talkers were made within the frequency band of 900 Hz (where intelligibility was M = 50.77 RAU) and the higher filter of 1100 Hz did not provide any additional cues. It can be concluded that a wider filter of 900 Hz is necessary to obtain sufficient amount of information about the dialect of female talkers and, possibly, a slightly higher amount of verbal cues (relative to males) is helpful in making decisions about the dialect. Possibly, for female speech, it is the 900 Hz band (and not 700 Hz) that is the optimal filter for removing the semantic content while

retaining most of the dialect-related cues. This male vs. female discrepancy within the 700 Hz – 900 Hz frequency region will need to be tested separately in a more focused design.

References

Arvaniti, A., & Garding, G. (2007). Dialectal variation in the rising accents of American English. In Cole, J., and Hualde, J. (eds), Laboratory Phonology 9, Mouton de Gruyter, Berlin, pp. 547–576.

Bezooijen, R. & Gooskens, C. (1999). Identification of language varieties: The contribution of different linguistic levels, Journal of Language and Social Psychology 18, 31-48.

Burnham, D., Kitamura, C., & Vollmer-Conna, U., 2002. What's new pussycat? On talking to babies and animals. Science 296, 1095.

Clopper, C., & Pisoni, D. (2004). Homebodies and army brats: Some effects of early linguistic experience and residential history on dialect categorization. Language Variation and Change 16, 31-48.

Clopper, C., Pisoni, D., & de Jong, K. (2005). Acoustic characteristics of the vowel systems of sic regional varieties of American English. Journal of the Acoustical Society of America 118, 1661-1676.

Clopper, C., & Pisoni, D. (2007). Free classification of regional dialects of American English. Journal of Phonetics 35, 421-438.

Clopper, C., & Smiljanic, R. (2011). Effects of gender and regional dialect on prosodic patterns in American English. Journal of Phonetics 39, 237-245.

Clopper, C., & Smiljanic, R. (2015). Regional variation in temporal organization in American English. Journal of Phonetics 49, 1-15.

Donaldson, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General, 121*, 275-277.

Fox, R., & Jacewicz, E. (2009). Cross-dialectal variation in formant dynamics of American English vowels. Journal of the Acoustical Society of America 126, 2603–2618.

French, N.R., & Steinberg, J.C., 1947. Factors governing the intelligibility ofspeech sounds. J. Acoust. Soc. Amer. 19, 90–119.

Frota, S., Vigario, M., & Martins, F. (2002). Language discrimination and rhythm classes: Evidence from Portuguese. In Proceedings of Speech Prosody 2002, Aix en Provence, pp. 319-322.

Green, D. M., & Swets. J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Irons, T. L. (2007). On the status of low back vowels in Kentucky English: More evidence of merger. Language Variation and Change 19, 137-180.

Jacewicz, E., Fox, R. A., O'Neill, C., & Salmons, J. (2009). Articulation rate across dialect, age, and gender. *Language Variation and Change*, 21, 233-256.

Jacewicz, E., Fox, R., & Wei, L. (2010). Between-speaker and within-speaker variation in speech tempo of American English. Journal of the Acoustical Society of America 128, 839-850.

Jacewicz, E. & Fox, R. A. (2012). The effects of cross-generational and cross-dialectal variation on vowel identification and classification. Journal of the Acoustical Society of America 131, 1413-1433.

Jacewicz, E. & Fox, R. A. (2014). The effects of indexical and phonetic variation on vowel perception in typically developing 9- to 12-year-old children. Journal of Speech, Language, and Hearing Research 57, 389-405.

Jacewicz, E., Fox, R. A., & Ortega, H. (2015). Source versus spectral cues in the perception of indexical features in speech. Journal of the Acoustical Society of America 137, 2417.

Kitamura, C., & Burnham, D., 2003. Pitch and communicative intent in mother's speech: adjustments for age and sex in the first year. Infancy, 4, 85–110.

Kitayama, S., & Ishii, K. (2002). Word and voice: Spontaneous attention to emotional utterances in two languages. Cognition and emotion, 16, 29-59.

Knoll, M. A., Uther, M., & Costall, A. (2009). Effects of low-pass filtering on the judgment of vocal affect in speech directed to infants, adults and foreigners. Speech Communication 51, 210-216.

Labov, W., Ash, S., & Boberg, C. (2006). Atlas of North American English: Phonetics, Phonology, and Sound Change. Mouton de Gruyter, Berlin.

Lass, N. J., Almerino, C. A., Jordan, L. F. & Walsh, J. M. (1980), The effect of filtered speech on speaker race and sex identications, Journal of Phonetics 8, 101-112.

Lynn, S. K., & Barrett, L. F. (2014). "Utilizing" signal detection theory. *Psychological Science, 25*, 1663-1673.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide (2nd ed)*. Mahwah, NJ: Erlbaum.

McNally R. J., Otto M. W., & Hornig C. D., "The voice of emotional memory: content-filtered speech in panic disorder, social phobia, and major depressive disorder." Behaviour research and therapy, vol. 39, no. 11, pp. 1329–37, Nov. 2001.

Pollack, I., 1948. Effects of high pass and low pass filtering on the intelligibility of speech in noise. J. Acoust. Soc. Amer. 20, 259–266.

Purnell, T., Salmons, J., & Tepeli, D. (2005). German substrate effects in Wisconsin English: Evidence for final fortition. American Speech 80, 135-164.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms, Speech Communication 40, 227–256

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117,* 34-50.

Studebaker, G. (1985). A "rationalized" arcsine transform. *Journal of Speech, Language, and Hearing Research, 28*, 455–462.